**The Examining Validity and Reliability of the Prosocial Behavior Scale: Insights from an Indonesian Sample**

**Abstract**

In the context of behavior, prosocial behavior is needed to maintain individual relationships with others. This behavior has existed among Islamic college students, who are known as Santri. Santri, as an adolescent, has the concept of establishing relationships through prosocial behavior. Therefore, a measuring instrument is needed to determine Santri's prosocial behavior. The Prosociality Scale is one of the popular instruments in measuring prosocial behavior, and it is essential to adapt and validate the instrument with the characteristics of the Indonesian sample, especially Santri. This study presents evidence of validity and reliability using the Rating Scale Rasch Measurement Model. With a sample of 742, of which 424 (57.1%) were female and 318 (42.9%) were male. This instrument has met the criteria of good psychometrics according to the Rasch model in terms of reliability (0.82 - 0.99) and validity and fulfills the need for measurement invariance.

Keywords:

*Prosocial Behavior, Adolescents, Rasch Model, Validation*

**Introduction**

In Indonesia, Islam has been acculturated and transformed into a majority religion that prioritises interfaith harmony (Anwar, 2018), as well as respect for differences in ethnicity, culture, and language, in order to maintain an intact social order (Maate, 2017). With adherents reaching 229.62 million people, or around 87.2% of Indonesia's total population (Mastuki, 2020), Islam provides guidance in education that prioritises religious values. This education is manifested in an institution called Pesantren. Pesantren is the oldest educational institution in Indonesia (Marzuki et al., 2020), which was adopted from the Hindu Buddhist educational tradition. Those who study in pesantren are called santri. In fact, the root of the word pesantren itself comes from the word santri, which is defined as a student, or earlier, Shastri, which is someone who has expertise in explaining the holy books of Hinduism (Islam & Aziz, 2020).

Santri cannot be separated from the pesantren tradition and is the most important part of the survival of the pesantren education pattern to this day (Dhofier, 2011). The reason is that the relationship between santri and pesantren cannot fade just like that after the santri finish their education in the pesantren. The relationship continues to be attached to the santri's lifetime. As a result, pesantren that hone santri to continue learning and provide benefits to society will continue to be a reference for santri behavior throughout their lives (Baso, 2012).

In terms of psychological characteristics, santri show a higher meaning of life, life optimism, and prosocial behavior than those who are not santri. Santri are also able to adapt to any condition, thinking that the suffering that comes to them is considered wisdom to continue to be optimistic in looking at life (Nashori, 2011). Santri's character cannot be separated from their participation in the independence of the Republic of Indonesia. One of them is the formation of Laskar Hizbullah in 1943, when Santri became part of the Laskar. And now, Laskar Hizbullah and Laskar Pembela Tanah Air are the forerunners of BKR and TKR, which are the first military institutions in Indonesia (Zuhri, 2013). The attitude of defending the homeland and sacrificing and prioritising the interests of the community selflessly is one of the characteristics of social behavior (Wittek & Bekkers, 2015).

Prosocial behavior in Santri cannot be separated from Islamic principles. Islam as a religion emphasises the teaching to help and provide assistance to others without expecting anything in return (Abdel-Khalek, 2013), more generally, Islam teaches compassion (Hanafi, 2001). herefore, santi behavior is very close to social behavior (Iffan, 2019). The concept of prosocial behavior covers a wide range of things, not limited to helping, sharing, entertaining, contributing, or offering services to someone else, but all things that are intended to benefit others. The study of prosocial behavior not only examines the factors that contribute to the behavior but also looks at its impact (Dovidio & Banfield, 2015). Meanwhile, measurements of prosocial behavior in Santri are not found in Indonesia or other countries. If it is not said to be a measurement, only Rosset (2016) research wants to explain the differences in the prosocial behavior of antri. Even then, it only relies on the t-test in its analysis. In fact, measuring the prosocial behavior of santri is important because prosocial behavior is a key aspect of the social and moral development of santri (Sabiq, 2012).

Santri, as students in pesantren, or religious education institutions, must have positive characters, morals, and values within themselves. In this case, pesantren, or religious education institutions, have an obligation to form individuals with noble character and integrity (Huda et al., 2023). Prosocial behaviors, such as empathy, caring, cooperation, and helping others, are an important part of forming good character and strong morals (Baldassarri & Abascal, 2020). In addition, prosocial behavior can help Santri understand the importance of contributing to society and the surrounding environment. It teaches one to be an active and positive member of the community and helps in building good relationships with others. Prosocial behavior teaches the importance of listening, respecting the views of others, and helping in situations that require the individual's own support (Barr & Higgins-D’Alessandro, 2007).

In its development, prosocial behavior in Santri can prepare them for a better adulthood. For example, the social and emotional skills gained through prosocial behavior will help them overcome social challenges and cooperate in various environments later in life (Akelaitis & Lisinskiene, 2018; Brownell, 2013; Kaltwasser et al., 2017). Prosocial behavior in Santri can also help them build understanding and tolerance for differences and diversity and promote inclusiveness in their social environment (Baldassarri & Abascal, 2020). Due to the importance of prosocial behavior in Santri, it is expected that they get holistic self-development so that they can grow into balanced and quality individuals.

Santri in Indonesia today are adolescents, whose development is strongly influenced by their surroundings and try to influence their environment. Prosocial behavior in adolescents, defined as voluntary actions meant to benefit others (Eisenberg et al., 2006) has been associated with several kinds of positive results, such as strong interpersonal bonds, academic achievement, and high self-esteem (Laible et al., 2004; Padilla-Walker & Carlo, 2014; Wentzel, 1993). Few longitudinal studies have looked at changes in prosocial behavior across a broad age range in adolescence, despite the fact that prosocial development has long been studied and general age-related increases have been reported from infancy through early adulthood (Crocetti & Rubini, 2017; Eisenberg & Fabes, 1998). Examples of these studies include Carlo (2015) and Luengo Kanacri (2013).

However, adolescence brings about a variety of behavioral, cognitive, and physical changes that affect social functioning. First, adolescents may be able to participate in a greater range of prosocial activities due to their increased physical maturity and level of autonomy (Fabes et al., 1999; Padilla-Walker & Carlo, 2014). Second, improvements in perspective taking (Van der Graaff et al., 2014, for example) might help people engage in prosocial conduct by facilitating higher-order moral thinking (Blasi, 1980; Eisenberg et al., 2006). Third, as social competence rises, so does the frequency of peer interactions and interest in close, romantic relationships (Steinberg & Morris, 2001).

At this stage, measurement instruments for Santri as adolescents is warranted. One of the reputable instruments for measuring prosocial behavior is the APBAScale developed by Caprara et al (2005). The instrument certainly be translated into Indonesian. Given the increasing need for an accurate and reliable instrument as well as the unavoidable diversity of cultures, languages, and races, instrument adaptation is common. Like the adult prosociality behavior instrument, which aims to integrate Indonesian conditions with Italian culture despite the instrument's Italian origins and development. However, the necessity for instrument adaptation cannot be guaranteed to produce accurate and reliable results. In order to create two equal instruments, the proper procedures must be followed, one of which is adhering to the International Test Commission's (ITC) recommendations for the cross-cultural translation and adaptation of psychological instruments.

In addition, the Indonesian version of the Prosociality Scale was determined to be trustworthy and to meet the goodness of fit index criteria after being modified and subjected to confirmatory factor analysis using a sample of university students (Sefianmi et al., 2023). However, we found that there was no explanation of the method of testing the results of the adaptation of the measuring instrument. Sefianmi et al (2023) only lists the use of the confirmatory factor analysis method without explaining the shortcomings of the method when used on categorical or ordinal data. As a result, bias will occur in the resulting score because confirmatory factor analysis treats categorical data as if it were a continuum (Cai, 2010).

One psychometric method that is qualified to test the validity and reliability of a scale or instrument is the Rasch Mode (Rasch, 1960). The advantage of this method is when individuals and items can be compared on the same straight line on the logit scale (Andrich & Marais, 2019). Thus, individuals and items can be compared starting from the difficulty level of the item and the individual's ability to respond to the item (Hayat et al., 2021), as well as the concept of objective specification, where item calibration is not tied to the person and the estimated person score is not tied to the item (Bond & Fox, 2015).

Thus, the existence of good methods in testing instruments and the importance of measuring prosocial behavior in Santri not only provide individual benefits but also encourage the formation of a better society as a whole.

**Literature Review**

**Defining and measuring Prosocial behavior**

Prosocial behavior is any action or pattern of behavior that seeks to improve the welfare of others or the general population. In positive psychology, prosocial activity is seen as the polar opposite of hostile behavior (Malonda et al., 2019). This is due to the fact that aggressive conduct is distinguished by behaviors that are harmful, hostile, and self-serving, whereas prosocial behavior comprises actions that are beneficial, cooperative, and caring towards others.

Furthermore, prosocial action is typically described as voluntary, implying that people engage in it of their own free will and without any outside pressure or force (Lam, 2012). Prosocial behavior is motivated by a real desire to assist others rather than by self-interest, the expectation of reward, or public recognition. Although there isn't a single definition that is accepted worldwide, prosocial behavior is typically understood to be behavior that is intended to help others without expecting anything in return or for one's own benefit. One can see these behaviors in a variety of contexts, such as those of families, friendships, communities, and even society at large (Eisenberg et al., 2010).

One instrument that measures prosocial behavior is The Prosociality Scale developed by Caprara (2005) among Italian sample. Each of the four dimensions of the Prosociality scale will have four items, making a total of 16 items in this instrument. The Prosociality Scale is a widely used scale for measuring adult prosociality. The Prosociality Scale is a self-report measure that assesses an individual's prosocial behavior. Participants rate the truth of each statement on a five-point Likert scale using the following options: never/nearly never true, rarely true, sometimes true, frequently true, and almost always/always true for each of the measure's 16 items. Without concentrating on particular prosocial behavior patterns, the scale evaluates adults' general prosocial behavior. The scale's components are intended to capture various prosocial behaviors, including empathy and sympathetic reactions as well as sharing, helping, and caring behaviors (Martí-Vilar et al., 2020).

For several reasons, The Prosociality Scale differs from other prosocial measures. The Prosociality Scale, as we know, evaluates prosocial behavior in adults, while additional measures may be customised to children or adolescents (Zhan et al., 2023). Furthermore, The Prosociality Scale evaluates prosocial behavior in general rather than emphasising specific prosocial behavior classifications (Luengo Kanacri et al., 2021; Martí-Vilar et al., 2020), and the overall The Prosociality Scale has strong reliability and validity and has been used in numerous studies across various countries (Luengo Kanacri et al., 2021; Zhan et al., 2023). There is some proof that the take care of Scale's usefulness varies among cultures. For instance, a research investigation conducted in Italy and Spain discovered that the Spanish version of the scale showed appropriate model fit in both the Spanish and Italian groups, and the results duplicated the instrument's adequacy reported in the Italian sample (Martínez-Pampliega et al., 2018).

**Methods**

**Procedure**

An online survey was conducted using a Google form, to which 742 santri responded. According to APA guidelines on research ethics, the researcher adhered to ethical standards when using the non-probability sampling technique. Before the poll was performed, all respondents were made aware that their participation was completely voluntary. Additionally, if they believed there was an irregularity in the survey administration, respondents had the option of stopping the survey in the process. Respondents received notice that any personal information they unintentionally provided in the course of the survey would be kept private, and all identifying information that may be used to link them to particular personal data was anonymized.

**Respondents**

Several categories were created from the 742 santri that completed the poll. 424 (57.1%) santri from the gender category were female as santriawati, and 318 (42.9%) were male as santriawan. The mean age of the study sample was 16.96 (SD = 1.34).

**Measurement**

In this study, the Prosociality Scale—a tool for measuring prosociality—was the object of validation. This 16-item scale was created by Caprara (2005), and includes three dimensions—helping, sharing, taking care of others . While the empathy component is still up for debate as to whether it should be included in the prosocial framework, the first three dimensions are characteristic of the measuring of prosocial conduct. Due to the fact that an individual's empathic motivations or dispositions are not only a correlate of their desire to act prosocially, but rather a fundamental component of that tendency (Eisenberg & Fabes, 1998) this is the case. But it's not implausible that a number of components may eventually come together to build the framework of empathy. As we know, Capara's prosocial behavior has been updated, so we still use four dimensions to further ensure prosocial behavior in accordance with Eisenberg's (1995) theory, which are helping, sharing, taking care of others (caring) and empathy. One point (strongly disagree) to four points (strongly agree) are used on a Likert-type scale for each topic.

**Translation Procedure**

The Indonesian Prosociality Scale refers to the guidelines of the International Test Commission regarding the translation and adaptation procedures of the test. First, two researchers who have an understanding of and fluency in English and Indonesia have translated the Prosociality Scale into Indonesian. Then, the translation results were discussed through a group discussion forum to discuss misunderstandings in translation, and a number of language modifications were made at this stage. Secondly, translations into English were done back from the result to compare with the original version. The re-translation is done by professional linguists and psychologists. Finally, a discussion group forum was held to evaluate factor and items that had been translated into Indonesian. The reviewers involved, ranging from forum members to translators, are holders of doctoral degrees and professionals from a number of universities located in Indonesia. From the result of the translation, the Indonesian Prosociality scale has a total of 16 items, which are divided into four items to helping (item 1, 3, 6 and 7), four to sharing (item 2, 9, 11 and 14), four to caring (item 4, 10, 13, and 14), and four to empathy (item 5, 8, 12, and 16)

**Statistical analysis**

***Confirmatory Factor Analysis***

The CFA model ascertains the relationships between the observable and latent variables. One kind of structural equation model is CFA, which offers an effective way to test various hypotheses regarding a number of measurable variables (Flora & Curran, 2004). n this study, we want to analyze the structural validity of the prosocial behavior instrument, which consists of four dimensions, namely, helping, sharing, taking care of others (caring), and empathy. The first test was conducted by establishing a unidimensional model, then testing with four dimensions and testing with a higher-order factor model. The analysis used Mplus 8.4 software (Muthen & Muthen, 2017), where the estimators we use are the weighted least squares mean estimator and the variance corrected estimator.

This study employed a number of goodness-of-fit indices to evaluate the adequacy of the proposed models. The individual model was assessed using the chi-square test, and the degree to which the model did not fit well in comparison to a perfect model was estimated using the standardized root mean squared residual (SRMR) and the root mean square error of approximation (RMSEA) (Browne & Cudeck, 1993). reater misspecification is associated with larger values; an RMSEA value of less than 0.05 is thought to be indicative of a model that fits adequately, and smaller SRMR values are linked to models that fit better; scores of less than 0.05 are thought to be indicative of a good fit (Maydeu-Olivares & Joe, 2014). Lastly, the comparative fit index (CFI) (Bentler, 1990) was also used as an incremental fit index index. The fit between the given model and a null model could be compared thanks to the CFI. A CFI score of 0.90 is typically seen as indicative of an appropriate model (Hu & Bentler, 1999).

***The Rasch Model***

The Rasch Model (Rasch, 1960) is a model that determines the probability of the response to each item based on the individual's level of the latent construct (i.e., the santri's prosocial level) and the item's level of difficulty. The model can be described as:

$$P\left(X\_{pi}=1\right)=\frac{e^{(β\_{p}-δ\_{i})}}{1+e^{(β\_{p}-δ\_{i})}}$$

Where $β\_{p}$ denotes the prosocial latent concept level and $δ\_{i} $denotes the item difficulty level. Therefore, it may be said that if students can exhibit prosocial activity, $δ\_{i} $provides an estimate of the rise in prosocial behavior. Initially merely dichotomous, the Rasch Model undergone development into a polytomous model, which started with the Rating Scale Model (Andrich, 1978). The function of the rating scale model was as follows:

$$log (P\_{nik}/P\_{ni (k-1)}) = B\_{n}+D\_{i}+F\_{k}$$

$P\_{ni (k-1)} $is the probability for *n* person to select the (*k - 1*) category, and $P\_{nik} $is the likelihood that *n* person would complete item *i* in category *k*. In the meantime, $D\_{i} $denotes the item's degree of difficulty (assuming that, in the statement of agreement, the respondent's difficulty is the difficulty of agreeing with the statement), and $B\_{n}$ level denotes the ability of the n-th individual. The likelihood that category *k* will be chosen based on category *k - 1* is thus represented by $F\_{k}$. A logit scale or log odds ratio is used to express $D\_{i} $and $B\_{n}$ (Linacre, 2006).

Because the RSM matches the instrument with the Likert format and the identical response category for every item, it will be used in this investigation. RSM will identify the item's placement at the difficulty level and the person's location at the trait level in its analysis (Wright & Masters, 1982). In addition, RSM still has assumptions that must be met, specifically, unidimensionality, parallel item characteristic curves, and local independence (Mair, 2018). Based on the Rating Scale Model, validity tests were carried out using Winteps 3.63.

**Result and Discussion**

**Confirmatory Factor Analysis**

Using the entire sample, the fit indices of the four-dimensional, higher-order factor, and unidimensional models were compared. The CFA results showed that the unidimensional model reasonably fit the data [χ2 (90) = 202.281, p < 0.001; RMSEA = 0.041 (90% CI = 0.033 -0.049), CFI = 0.941, SRMR = 0.052], compared to four dimensions model [χ2 (97) = 474.157, p < 0.001; RMSEA = 0.072 (90% CI = 0.066 -0.079), CFI = 0.802, SRMR = 0.072] and the higher order factor model [χ2 (100) = 520.836, p < 0.001; RMSEA = 0.075 (90% CI = 0.069  0.082), CFI = 0.779, SRMR = 0.074]. Hence, this result indicates that a unidimensional model is more appropriate for this study sample. All items have significant loadings in the range of 0.147 - 0.602, of which item number 1 needs special attention due to its small factor loadings.

**Dimensionality**

PCAR was utilized to examine the measuring instrument's unidimensionality assumption (Chou & Wang, 2010; Smith, 2002). The prosocial instrument's measurement model turned out to be unidimensional. As a result, the analysis's findings verified that the unidimensionality requirement of the Rasch RSM had been satisfied and that more research was warranted. Based on PCAR, a test is considered to measure a dimension only if the measure's minimal variance explained is greater than 30% (Linacre, 1998). The test used here demonstrated unidimensionality, as evidenced by values greater than 38.5% (16.0 in eigenvalues unit) of the variance explained by the measure. This conclusion is consistent with the findings regarding the factor structure after the CFA analysis.

**Local Independence**

Local independence is the basis of the Rasch RSM. For an examinee or examinees at a certain competence level, local independence refers to the ability of one item's performance to stand alone from another item's performance (Mair, 2018). The assumption of local independence was examined using the Q3 statistic (Yen, 1984) once it was established that the unidimensionality assumption had been satisfied. No items showed signs of local dependence when analyzed using the Q3 statistic index criteria, which state that the raw residual correlation between pairs of items is never > 0.30 (Christensen et al., 2017; Das Nair et al., 2011). Items 9 and 11, both at 0.23, which is less than 0.30, had the highest raw residual correlations. From the report as mentioned above, the research assumption of local independence has been met.

**Item Fit**

To ascertain how well each item helps to create a single common construct as proof of scale unidimensionality, item fit metrics like infit and outfit MNSQ statistics can also be adopted. Rasch RSM requirements say that an outfit or infit MNSQ value of 1 is desirable, while values between 0.5 and 1.5 work well for measurement (Andrich & Marais, 2019; Bond & Fox, 2015). Through the use of infit and outfit MNSQ data, it was determined that all 16 items on the prosocial instrument were within the permissible range of 0.5–1.5. Furthermore, the prosocial instrument's point measure correlation range, as presented in Table 1, was 0.43 to 0.61, meaning that every item had changed in the same direction (Bond & Fox, 2015). This outcome provided more support for the Rasch RSM's conclusions in this investigation. Rasch item-fit statistics generally confirmed the unidimensionality of the prosocial instrument scale.

**Rating Scale Diagnostic**

In the step order description (see Table 2), it is revealed that the average participant who chooses order one on this scale is estimated to be –3.14, which means that the average ability increases as the order increases or can be interpreted from negative to positive. In addition, the threshold between points 1 and 2 is -1.82, for points 2 and 3 is –0.47, and for points 3 and 4 is 2.36. So, the scale threshold increases as each scale point increases, and also from negative to positive. There was no step disorder on this scale, indicating that the four-point scale used in the prosociality scale was functioning at the same response level as intended by the test developers. The category answer function of the prosocial instrument is represented graphically in Figure 1 by the results of the scale analysis. The suggested pattern is depicted in the graph, where the most likely response for each ability level at each scale competency is linked. Overall, the analysis showed that the rating scale performed as intended.

**Table 1.** Item Fit Measure

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Item | Measure | SE | Infit | Outfit | Rpm |
| Item 1 | 1.55 | 0.05 | 1.28 | 1.31 | 0.43 |
| Item 2 | -0.05 | 0.06 | 0.77 | 0.76 | 0.60 |
| Item 3 | 0.48 | 0.06 | 0.96 | 0.94 | 0.51 |
| Item 4 | 1.29 | 0.06 | 1.01 | 1.04 | 0.50 |
| Item 5 | 0.81 | 0.06 | 1.06 | 1.10 | 0.53 |
| Item 6 | 0.67 | 0.06 | 0.92 | 0.94 | 0.57 |
| Item 7 | 0.35 | 0.06 | 0.93 | 0.93 | 0.52 |
| Item 8 | 0.05 | 0.06 | 0.89 | 0.87 | 0.61 |
| Item 9 | -0.46 | 0.06 | 0.94 | 0.91 | 0.52 |
| Item 10 | -1.35 | 0.07 | 1.30 | 1.22 | 0.43 |
| Item 11 | -0.93 | 0.07 | 0.91 | 0.84 | 0.53 |
| Item 12 | 0.36 | 0.06 | 1.15 | 1.15 | 0.53 |
| Item 13 | -0.29 | 0.06 | 0.97 | 0.94 | 0.55 |
| Item 14 | -0.83 | 0.07 | 1.05 | 1.00 | 0.52 |
| Item 15 | -1.20 | 0.07 | 1.05 | 0.98 | 0.50 |
| Item 16 | -0.44 | 0.06 | 1.03 | 0.99 | 0.54 |

**Table 2.** Category Stucture

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Responses | Category Score | Measure | Infit MNSQ | Outfit MNSQ | Andrich threshold |
| Strongly Disagree | 1 | -3.14 | 1.21 | 1.3 | - |
| Disagree | 2 | -1.21 | 0.95 | 0.95 | -1.89 |
| Agree | 3 | 1.00 | 0.94 | 0.89 | -0.47 |
| Strongly Agree | 4 | 3.50 | 1.01 | 0.99 | 2.36 |

*Infit MNSQ: information-weighted mean square statistic; Outfit MNSQ: outlier-sensitive mean square statistics*

**Figure 1.** Category response curve of all Prosocial items



**Reliability**

Reliability is calculated in the Rasch RSM for both individuals and objects. For the prosocial instrument, the person separation reliability (a measure of how successfully an instrument separates people based on the variable being measured) was 0.82 (Malec et al., 2007; Wright & Masters, 1982). Thus, 2.16, expressed in standard error units, was the person separation index used to estimate the distribution of people on the measured variable. This value suggested good inter personal separation. Excellent psychometric features were shown by the Prosocial instrument, with reliability and separation for items calculated in the same way as for individuals, coming in at 0.99 and 12.80, respectively. with the provision of person reliability above 0.80 and item reliability above 0.90 (Malec et al., 2007). From this value, the consistency of person and item on the Indonesian Prosocial Scale is good.

**Wright map**

The distribution of people and items as shown by a Wright map. Information on the comparison between person and item locations is displayed in the visualization in Figure 1. Each # gives a description of 10 persons, and "." is a range of 1 to 9 persons, with the ability location on the left containing 742 participants and the difficulty location containing 16 items.

Without relying on one another's estimates, the Rasch model offers individual estimates that can be compared to the results of calibrating item difficulty in logit units. The comparison shows that the item's mean -which in the Rasch model is always 0- is lower than the individual's mean, which may signify that the person is more adept at or in agreement with the prosocial scale than the item's response difficulty. Each item has tough and easy values, and no size is lost too far when compared to a meter or ruler between objects. There is also no significant gap between items in their logit location. The prosociality scale's 16 items, however, fall short of the required proficiency for measures two logit and higher.



*Key*:

<more> = most able persons

<less> = least able persons

<rare> = most difficult to endorse items

(freq> = least difficult to endorse items

# = 10 person

. = 1 to 9 person

M = the location mean persons and mean items

S = one standard deviation away from the mean

T = two standard deviation away from the mean

**Figure 2.** Item and Person Map



**Figure 3.** Test information function of Prosocial

**Test Information Function**

Finding the point on the agreement scale where a test yields the most accurate estimate of prosocial agreement is of great interest once the pertinent data for each item has been identified. The test information function (TIF) θ is only the total of the information functions for each item. Figure 3 below shows the TIF for the scale. A graph pertaining to the prosocial instrument's measurement information function is displayed in Figure 3. The results of the prosocial instrument showed that the relatively moderate opportunity level for agreement had the highest information function, and a significant amount of information was gleaned from the data. Both the latent trait level and the information derived from the measurements were relatively low at a high opportunity level for agreement. These findings demonstrated that when the prosocial instrument was administered to those with intermediate trait levels, the best information was obtained.

**Differential item functioning (DIF) testing for invariance**

The Rasch model's measurement invariance creates circumstances in which the calibration items are unaffected by person influence, and the expected person scores are unaffected by item influence (Wright, 2006). In essence, no single factor may lead to the specific assumption of objectivity being violated. As a result, the Rasch model, in this case, also includes DIF criteria to find the presence of objects with various functions. To ascertain the possibility of items being exposed to DIF, the reference and focal groups' scores are compared. The Mantel-Haenszel test and the Rasch-Welch t-test are the two methods for testing DIF in the Rasch model. On a scale, the latter approach is more accurate and sensitive at spotting DIF (Paek & Wilson, 2011). In reality, the logistic regression model is employed as a tool to estimate each individual or group, with the assumption that the degree of difficulty is constant or uniform across all groups. Differences in difficulty level ratings between groups on particular items are then used as a sign of DIF, allowing some groups to receive better opportunities than other groups. In general, researchers can evaluate changes in answer patterns where one group gains more than another if the significance of the difference is less than 0.05. Using effect size to calculate DIF is a different strategy. According to Zieky (1993), a researcher can assume that an item indicates DIF if the contrast difference across groups on that item is more than 0.64 (Song et al., 2020) and the Rasch-Welch t-test shows a value below 0.5 in probability. In this analysis, there are no items exposed to DIF; it is just that there are a number of items that are easier for males to answer, namely item 6, item 12, item 13, and item 16.

**Discussion**

This study aims to test the validity of a prosociality measurement tool with a sample of santri as the reference. Santri as adolescents are considered as agents in modeling moderate behavior as well as behavior that supports social morals. So it is very important to know how the prosocial attitude of santri. The Indonesian Prosociality Scale has undergone several stages of validity testing, and is certainly different from its original structure. By using Rasch Model-based Item Response Theory, the evidence provided will be stronger and more rigorous regarding the results of the validity testing of the measuring instrument. Thus, from this study, we gained some insights.

First, The development of a four-dimensional model forming the latent prosocial variable proposed by (Caprara et al., 2005) did not support this study. After preliminary analysis of the factor structure of the prosocial scale, this research sample refers to a unidimensional model on 16 items measuring one prosocial factor. Changes in factor structure are possible in any measurement of latent variables. Especially in relation to the characteristics of the sample used, because it contains cultural, linguistic, ethnic and other diversity (Neumann et al., 2008).

Second, the Rasch model provides evidence of the validity of the measuring instrument content using reference to the Infit and Outfit scores. These two scores are expected to evaluate the validity of the item in terms of its content. If the item exceeds the limit of the specified criteria, then the item must be removed, not the model of the calculation is changed, this is what makes the Rasch model a good tool in validity testing. All of item The Indonesian Prosociality Scale has passed the Infit and Outfit evaluation in the Rasch model. But, a note that certain items require review from experts, in this case item 1 helping ("I am pleased to help my friends/colleagues in their activities "), and this is in line with the results of a low factor loading on confirmatory factor analysis that considers this item. In the Islamic boarding school environment, the words "help" and "pleased" sound strange, because the culture in Islamic boarding schools makes it mandatory to help someone, especially with kindness. Adolescents who are basically students find it difficult to accept this. The attitude of helping has become a habit because of their communal life in the context of the learning place.

Third, the expansion is obtained from the description of the person and item map, which explains the comparison of person and item on the same linear line. With logit units, The Indonesian Prosociality Scale has items that are quite easy for santri, namely with the logit score of santri above the logit value of the item. In this finding, the items of The Indonesian Prosociality Scale are actually not suitable for seeing prosocial attitudes in santri. The reason is, these items do not accurately describe the high prosocial attitude of santri considering that the items are quite easy to agree with. So, more precisely these items are intended to measure low individual prosocial attitudes. There needs to be development in this measuring instrument by making modifications by adding items to get the appropriate range in describing santri with high prosocial attitudes. Santri are known to hold strong norms related to religion. Religion emphasizes humans to have attitudes that are in accordance with social morals and support the good of society. Not only that, the pattern of education in pesantren also requires santri to prioritize social interests, and put aside personal interests, and help without any strings attached (Nahdiyah, 2018).

Fourth, the Indonesian Prosociality Scale has a good function of response order. There were no disordered step responses on the scale. This indicates something normal and the function of the scale corresponds to the adaptation of the test. With the evidence of the absence of step disorder, it shows that the pattern of the respondents' answers is normal. And conversely, if step disorder occurs, then it should be noted that there is an indication that the item has been misunderstood to mean the opposite of the latent variable.

Fifth, and the last is the invariance of the measuring instrument which is a reference to the balance of items in measuring prosocial. In this study, gender is the object to see the fulfillment of invariance on the scale. On the invariance measurement, it can be concluded that all items are not exposed to gender bias. At least there are some items that are easier on the other gender, for example, more favourable to women or men. However, it does not have a serious impact that causes DIF. Sometimes in terms of gender, for example women, we can give a picture that women have higher emotional support for colleagues than men, as well as sharing behavior that is above men (Nielson et al., 2017), so it is not impossible for women to answer the item more easily. Moreover, the conditions of pesantren that have nuances of independence and long-distance relationships with parents can cause a lack of feelings of affection. Women will look for friends who are able to relieve their problems, by prioritizing mutual sharing and helping each other (Martínez-Gregorio et al., 2023).

**Conclusion**

The Indonesian Prosociality Scale has been psychometrically validated on a sample of santri, and has good validity, although in this study there were changes to the factor structure that did not follow the original scale. This study illustrates that the Santri in the sample already have high prosocial behavior. This is possible due to the conditions of pesantren education that teach Santri to be kind in their social environment. As for the adaptation carried out with prosocial instruments developed outside Indonesia, it has yet to provide a good range in seeing the overall picture of students' prosocial behavior.

Then, we also believe that there are a number of limitations in this study. First, this study still does not ensure external validity due to data limitations. Although it has been passed with a number of validity tests, external validity is considered by the researcher to be important to be carried out in future studies. External validity can be done by examining the relationship of this measuring instrument with other measuring instruments that have the same concept, by examining the relationship between variables that measure prosocial behavior.

Second, in this study, DIF analysis was limited to gender, although it is known that many demographic variables need to be tested and interpreted more deeply. DIF testing is important to do from various sides, in order to ensure that the measuring instrument meets the needs of invariance.

Third, the addition of large-scale samples needs to be done to see the difference in the possibility of changes in the dimensional structure. As well as non santri samples are important to be re-tested and compared, given the large differences in the lifestyles of santri and non santri. The Indonesian Prosociality Scale has a number of research constraints that go beyond this study's findings. Future researchers must revisit these limitations and apply them to the prosocial behavior of teenagers or santri.

**Declarations**

**Confict of interest** The authors declare no competing interests.

**Author’s Contributions**

**Wahyu Syahputra** research concept and design, collection and/or assembly of data, data analysis and interpretation, writing the article, final approval of the article. **Ika Widhiastuti** collection and/or assembly of data, interpretation, writing the article. **Baydhowi** supervision, writing the article, final approval of the article. **Saiful Falah** supervision, writing the article, final approval of the article. **Devie Yundianto** data analysis and interpretation, writing the article. **Moondore Madalina Ali** supervision, writing the article, final approval of the article.

**Consent for publication**

Not applicable.

**Funding**

There is no source of funding.

**Data availability**

We guarantee the transparency of the data described in this manuscript. Data can be accessed on <https://zenodo.org/doi/10.5281/zenodo.13831300>

**Reference**

Abdel-Khalek, A. M. (2013). Personality dimensions and religiosity among Kuwaiti Muslim college students. *Personality and Individual Differences*, *54*(2), 149–152. https://doi.org/https://doi.org/10.1016/j.paid.2012.08.004

AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. American Educational Research Association. https://cir.nii.ac.jp/crid/1130000794987222016.bib?lang=en

Akelaitis, A. V., & Lisinskiene, A. R. (2018). Social emotional skills and prosocial behavior among 15-16-year-old adolescent. *European Journal of Contemporary Education*, *7*(1), 21–28. https://doi.org/10.13187/EJCED.2018.1.21

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*(4), 561–573. https://doi.org/10.1007/BF02293814

Andrich, D., & Marais, I. (2019). A Course in Rasch Measurement Theory. In *Springer Nature*. https://doi.org/10.1016/b978-008043348-6/50005-x

Anwar, C. (2018). Islam Dan Kebhinekaan di Indonesia: Peran Agama Dalam Merawat Perbedaan. *Zawiyah: Jurnal Pemikiran Islam*, *4*(2), 1–18. https://doi.org/10.31332/ZJPI.V4I2.1074

Aryadoust, V., Ng, L. Y., & Sayama, H. (2020). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing*, *38*(1), 6–40. https://doi.org/10.1177/0265532220927487

Badenes-Ribera, L., Duro-García, C., López-Ibáñez, C., Martí-Vilar, M., & Sánchez-Meca, J. (2022). The Adult Prosocialness Behavior Scale: A reliability generalization meta-analysis. *International Journal of Behavioral Development*, *47*(1), 59–71. https://doi.org/10.1177/01650254221128280

Baldassarri, D., & Abascal, M. (2020). Diversity and prosocial behavior. *Science*, *369*(6508), 1183–1187. https://doi.org/10.1126/science.abb2432

Barr, J. J., & Higgins-D’Alessandro, A. (2007). Adolescent Empathy and Prosocial Behavior in the Multidimensional Context of School Culture. *The Journal of Genetic Psychology*, *168*(3), 231–250. https://doi.org/10.3200/GNTP.168.3.231-250

Baso, A. (2012). *Pesantren Studies 2a*. Pustaka Afid.

Blasi, A. (1980). Bridging moral cognition and moral action: A critical review of the literature. *Psychological Bulletin*, *88*(1), 1.

Bond, T., & Fox, C. (2015). *Applying the Rasch Model. In Applying the Rasch model: Fundamental Measurement in the Human Sciences*. https://doi.org/10.1017/CBO9781107415324.004

Boone, W. J., Yale, M. S., & Staver, J. R. (2014). Rasch analysis in the human sciences. In *Rasch Analysis in the Human Sciences*. https://doi.org/10.1007/978-94-007-6857-4

Browne, M. W., & Cudeck, R. (1993). Alternative Ways of Assessing Model Fit. *Sociological Methods &amp; Research*, *21*(2), 230–258. https://econpapers.repec.org/RePEc:sae:somere:v:21:y:1992:i:2:p:230-258

Brownell, C. A. (2013). Early Development of Prosocial Behavior: Current Perspectives. *Infancy*, *18*(1), 1–9. https://doi.org/https://doi.org/10.1111/infa.12004

Cai, L. (2010). High-dimensional Exploratory Item Factor Analysis by A Metropolis–Hastings Robbins–Monro Algorithm. *Psychometrika*, *75*(1), 33–57. https://doi.org/10.1007/s11336-009-9136-x

Caprara, G. V., Steca, P., Zelli, A., & Capanna, C. (2005). A New Scale for Measuring Adults’ Prosocialness. *European Journal of Psychological Assessment*, *21*(2), 77–89. https://doi.org/10.1027/1015-5759.21.2.77

Carlo, G., Padilla-Walker, L. M., & Nielson, M. G. (2015). Longitudinal bidirectional relations between adolescents’ sympathy and prosocial behavior. *Developmental Psychology*, *51*(12), 1771.

Crocetti, E., & Rubini, M. (2017). Communicating Personal and Social Identity in Adolescence. *Oxford Research Encyclopedia of Communication*. https://doi.org/10.1093/acrefore/9780190228613.013.482

Davis, D. R., & Boone, W. (2021). Using Rasch analysis to evaluate the psychometric functioning of the other-directed, lighthearted, intellectual, and whimsical (OLIW) adult playfulness scale. *International Journal of Educational Research Open*, *2*, 100054. https://doi.org/https://doi.org/10.1016/j.ijedro.2021.100054

de Ayala, R. J. (2009). The Theory and Practice of Item Response Theory. In *International Journal of Testing*. https://doi.org/10.1080/15305058.2011.556771

Dhofier, Z. (2011). *Tradisi Pesantren : Studi Pandangan Hidup Kyai dan Visinya Mengenai Masa Depan Indonesia* (8th ed.). Pustaka LP3ES.

Dovidio, J. F., & Banfield, J. C. (2015). *Prosocial Behavior and Empathy* (J. D. B. T.-I. E. of the S. & B. S. (Second E. Wright (ed.); pp. 216–220). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-08-097086-8.24024-5

Eisenberg, N., Carlo, G., Murphy, B., & Van Court, P. (1995). Prosocial development in late adolescence: a longitudinal study. *Child Development*, *66*(4), 1179–1197.

Eisenberg, N., Eggum, N. D., & Di Giunta, L. (2010). Empathy-related Responding: Associations with Prosocial Behavior, Aggression, and Intergroup Relations. *Social Issues and Policy Review*, *4*(1), 143–180. https://doi.org/10.1111/j.1751-2409.2010.01020.x

Eisenberg, N., & Fabes, R. A. (1998). Prosocial Development. In *Handbook of child psychology: Social, emotional, and personality development, Vol. 3, 5th ed.* (pp. 701–778). John Wiley & Sons, Inc.

Eisenberg, N., Fabes, R. A., & Spinrad, T. L. (2006). Prosocial Development. In *Handbook of child psychology: Social, emotional, and personality development, Vol. 3, 6th ed.* (pp. 646–718). John Wiley & Sons, Inc.

Fabes, R. A., Carlo, G., Kupanoff, K., & Laible, D. (1999). Early adolescence and prosocial/moral behavior I: The role of individual processes. *The Journal of Early Adolescence*, *19*(1), 5–16.

Fisher, W. P. J. (2007). Rating Scale Instrument Quality Criteria. *Rasch Measurement Transaction*, 21–1095. https://www.rasch.org/rmt/rmt211m.htm

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor Analysis with Ordinal Indicators: A Monte Carlo Study Comparing DWLS and ULS Estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 625–641. https://doi.org/10.1080/10705510903203573

Hanafi, H. (2001). *Agama, Kekerasan dan Islam Kontemporer* (1st ed.). Jendela.

Hayat, B., Hidayat, R., & Putra, M. D. K. (2021). Exploring the factor structure of environmental attitudes measure in a sample of Indonesian college students. *Revista CES Psicologia*, *14*(1), 112–129. https://doi.org/10.21615/CESP.14.1.9

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huda, M. N., Duwila, M., Nurul, S., Mojokerto, I., Gowa, D., Pungging, K., & Mojokerto, K. (2023). Menantang Disintegrasi Moral di Era Revolusi Industri 4.0 : Peran Revolusioner Pondok Pesantren. *Journal of Islamic Education*, *9*(1), 1–13. https://doi.org/10.18860/JIE.V9I1.22805

Iffan, A. G. (2019). Santri dan Nasionalisme. *Islamic Insights Journal*, *1*(1), 41–45. https://doi.org/10.21776/UB.IIJ.2019.001.01.4

Islam, M. H., & Aziz, A. (2020). Transformation of Pesantren in Maintaining Good Character. *HUMANISTIKA : Jurnal Keislaman*, *6*(1), 35–48. https://doi.org/10.55210/humanistika.v6i1.307

Kaltwasser, L., Hildebrandt, A., Wilhelm, O., & Sommer, W. (2017). On the relationship of emotional abilities and prosocial behavior. *Evolution and Human Behavior*, *38*(3), 298–308. https://doi.org/https://doi.org/10.1016/j.evolhumbehav.2016.10.011

Laible, D. J., Carlo, G., & Roesch, S. C. (2004). Pathways to self-esteem in late adolescence: the role of parent and peer attachment, empathy, and social behaviors. *Journal of Adolescence*, *27*(6), 703–716. https://doi.org/10.1016/j.adolescence.2004.05.005

Lam, C. M. (2012). Prosocial involvement as a positive youth development construct: a conceptual review. *TheScientificWorldJournal*, *2012*, 769158. https://doi.org/10.1100/2012/769158

Linacre, J. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Meas Trans*, *16*.

Linacre, J. (2006). A User’s Guide to WINSTEPS/MINISTEP Rasch-Model Computer Programs (3.91.0). In *Chicago*. http://www.winsteps.com/a/Winsteps-ManualPDF.zip

Luengo Kanacri, B. P., Eisenberg, N., Tramontano, C., Zuffiano, A., Caprara, M. G., Regner, E., Zhu, L., Pastorelli, C., & Caprara, G. V. (2021). Measuring Prosocial Behaviors: Psychometric Properties and Cross-National Validation of the Prosociality Scale in Five Countries. *Frontiers in Psychology*, *12*, 693174. https://doi.org/10.3389/fpsyg.2021.693174

Luengo Kanacri, B. P., Pastorelli, C., Eisenberg, N., Zuffianò, A., & Caprara, G. V. (2013). The development of prosociality from adolescence to early adulthood: The role of effortful control. *Journal of Personality*, *81*(3), 302–312.

Maate, S. (2017). Peran Islam dalam Dinamika Sosial, Budaya, dan Ekonomi di Indonesia. *SOSIORELIGIUS: JURNAL ILMIAH SOSIOLOGI AGAMA*, *2*(1). https://doi.org/10.24252/SOSIORELIGIUS.V2I1.5999

Mair, P. (2018). *Modern Psychometrics with R*. Springer International. https://doi.org/https://doi.org/10.1007/978-3-319-93177-7

Malec, J. F., Torsher, L. C., Dunn, W. F., Wiegmann, D. A., Arnold, J. J., Brown, D. A., & Phatak, V. (2007). The mayo high performance teamwork scale: reliability and validity for evaluating key crew resource management skills. *Simulation in Healthcare : Journal of the Society for Simulation in Healthcare*, *2*(1), 4–10. https://doi.org/10.1097/SIH.0b013e31802b68ee

Malonda, E., Llorca, A., Mesurado, B., Samper, P., & Mestre, M. V. (2019). Parents or Peers? Predictors of Prosocial Behavior and Aggression: A Longitudinal Study. *Frontiers in Psychology*, *10*, 2379. https://doi.org/10.3389/fpsyg.2019.02379

Martí-Vilar, M., Merino-Soto, C., & Rodriguez, L. M. (2020). Measurement Invariance of the Prosocial Behavior Scale in Three Hispanic Countries (Argentina, Spain, and Peru). *Frontiers in Psychology*, *11*, 29. https://doi.org/10.3389/fpsyg.2020.00029

Martínez-Gregorio, S., Tomás, J. M., & Oliver, A. (2023). A Psychometric Study of the Prosocial Behavior Scale: Differential Item Functioning by Gender. In *Behavioral Sciences* (Vol. 13, Issue 3). https://doi.org/10.3390/bs13030259

Martínez-Pampliega, A., Cormenzana, S., Sanz, M., Barni, D., Simon, J., Alomar, E., & Pérez, C. (2018). Metric Goodness of the Adult Prosocialness Scale. Comparative Study of Italy and Spain. *The Spanish Journal of Psychology*, *21*, E29. https://doi.org/10.1017/sjp.2018.26

Marzuki, M., Miftahuddin, M., & Murdiono, M. (2020). Multicultural Education in Salaf Pesantren and Prevention of Religious Radicalism in Indonesia. *Jurnal Cakrawala Pendidikan*, *39*(1), 12–25. https://doi.org/10.21831/cp.v39i1.22900

Mastuki. (2020). *Menjadi Muslim, Menjadi Indonesia (Kilas Balik Indonesia Menjadi Bangsa Muslim Terbesar)*. https://kemenag.go.id/opini/menjadi-muslim-menjadi-indonesia-kilas-balik-indonesia-menjadi-bangsa-muslim-terbesar-03w0yt

Messick, S. (1989). Validity. In *Educational measurement, 3rd ed.* (pp. 13–103). American Council on Education.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons’ responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, *63*(7), 737–745. https://doi.org/10.1016/j.jclinepi.2010.02.006

Muthen, L. K., & Muthen, B. (2017). MPlus, Statistical Analysis With Latent Variables user’s guide. In *Muthén & Muthén* (8th ed.). http://www.ncbi.nlm.nih.gov/pubmed/22210541

Nahdiyah, H. (2018). Menggugah Keberadaan Pesantren Untuk Membangun Nalar Pendidikan Berbasis Local Wisdom: Suatu Tela’ah Eksploratif dalam Perspektif Budaya. *At-Turas*, *4*(2), 208–226. https://doi.org/10.33650/at-turas.v4i2.333

Nashori, F. (2011). Kekuatan Karakter Santri. *Millah: Journal of Religious Studies*, *11*(1 SE-Articles), 203–219. https://doi.org/10.20885/millah.vol11.iss1.art10

Nielson, M. G., Padilla-Walker, L., & Holmes, E. K. (2017). How do men and women help? Validation of a multidimensional measure of prosocial behavior. *Journal of Adolescence*, *56*, 91–106. https://doi.org/10.1016/j.adolescence.2017.02.006

Orcan, F. (2018). Exploratory and Confirmatory Factor analysis: Which one to use first? *Egit. Psikol. Olcme Deger. Derg.*, *9*(4), 414–421. https://doi.org/10.21031/epod.394323

Padilla-Walker, L. M., & Carlo, G. (2014). The study of prosocial behavior. *Prosocial Development: A Multidimensional Approach*, *3*.

Paek, I., & Wilson, M. (2011). Formulating the Rasch Differential Item Functioning Model Under the Marginal Maximum Likelihood Estimation Context and Its Comparison With Mantel–Haenszel Procedure in Short Test and Small Sample Conditions. *Educational and Psychological Measurement*, *71*(6), 1023–1046. https://doi.org/10.1177/0013164411400734

Rasch, G. (Georg). (1960). *Probabilistic models for some intelligence and attainment tests*. 199.

Rosset, D. (2016). *Perbedaan Perilaku Prososial Santriwati Antara Pesantren Modern Dan Tradisional* [Muhammadiyah Malang University]. https://eprints.umm.ac.id/34407/1/jiptummpp-gdl-debrinaros-44062-1-debrina-t.pdf

Sabiq, Z. (2012). Kecerderdasan Emosi, Kecerdasan Spiritual dan Perilaku Prososial Santri Pondok Pesantren Nasyrul Ulum Pamekasan. *Persona:Jurnal Psikologi Indonesia*, *1*(2). https://doi.org/10.30996/PERSONA.V1I2.21

Sefianmi, D., Purnama, C. Y., & Setyadiredja, A. N. (2023). Adaptasi Skala Prosocialness for Adult Versi Bahasa Indonesia. *Jurnal Psikologi*. https://doi.org/http://dx.doi.org/10.35760/psi.2023.v16i1.7284

Song, H.-D., Hong, A. J., & Jo, Y. (2020). Psychometric Investigation of the Utrecht Work Engagement Scale-17 Using the Rasch Measurement Model. *Psychological Reports*, *124*(3), 1384–1411. https://doi.org/10.1177/0033294120922494

Steinberg, L., & Morris, A. S. (2001). Adolescent Development. *Journal of Cognitive Education and Psychology*, *2*(1), 55–87. https://doi.org/10.1891/194589501787383444

Van der Graaff, J., Branje, S., De Wied, M., Hawk, S., Van Lier, P., & Meeus, W. (2014). Perspective taking and empathic concern in adolescence: gender differences in developmental changes. *Developmental Psychology*, *50*(3), 881.

Wentzel, K. R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology*, *85*(2), 357.

Wittek, R., & Bekkers, R. (2015). *Altruism and Prosocial Behavior, Sociology of* (J. D. B. T.-I. E. of the S. & B. S. (Second E. Wright (ed.); pp. 579–583). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-08-097086-8.32158-4

Wolfe, E. W., & Smith, E. V. J. (2007). Instrument development tools and activities for measure validation using Rasch models: part II--validation activities. *Journal of Applied Measurement*, *8*(2), 204–234.

Wright, B. (2006). Local Dependency, Correlations and Principal Components. *Rasch Measurement Transactions*, 509–511. https://www.rasch.org/rmt/rmt103b.htm

Wright, B., & Masters, G. (1982). *Rating Scale Analysis*. MESA Press.

Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A.-R. (2015). Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication. *Journal of Caring Sciences*, *4*(2), 165–178. https://doi.org/10.15171/jcs.2015.017

Zhan, Q., Wang, S., Li, C., Li, M., Liu, D., Peng, W., Song, F., Shi, T., & Li, Y. (2023). Revision and validation of the prosocialness scale for adults (PSA) among chinese college students. *BMC Psychology*, *11*(1), 124. https://doi.org/10.1186/s40359-023-01124-3

Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In *Differential item functioning.* (pp. 337–347). Lawrence Erlbaum Associates, Inc. https://doi.org/10.1075/z.62.13kok

Zuhri, S. (2013). *Berangkat dari Pesantren* (1st ed.). LKiS Yogyakarta.